The crystal structure of pneumococcal surface antigen PsaA reveals a metal-binding site and a novel structure for a putative ABC-type binding protein

Michael C Lawrence[1]*, Patricia A Pilling[1], V Chandana Epa[1], Anne M Berry[2], A David Ogunniyi[2] and James C Paton[2]

Background: The surface protein PsaA of the pathogenic bacterium *Streptococcus pneumoniae* plays an essential role in its virulence. PsaA is a putative ATP-binding cassette-type (ABC-type) binding protein involved in the uptake of Mn^{2+} and possibly Zn^{2+} and is considered to be both a potential drug target and and a candidate vaccine component.

Results: The structure of PsaA has been determined to 2.0 Å resolution using X-ray crystallography and is the first structure obtained for an ABC-type binding protein from a Gram-positive organism. The protein consists of two $(\beta/\alpha)_4$ domains linked together by a single helix. A metal-binding site is formed in the domain interface by the sidechains of His67, His139, Glu205 and Asp280 and is occupied in the structure.

Conclusions: The structural topology of PsaA is fundamentally different from that of other ABC-type binding proteins determined thus far in that PsaA lacks the characteristic 'hinge peptides' involved in conformational change upon solute uptake and release. In our structure, the metal-binding site is probably occupied by Zn^{2+}. The site seems to be well conserved amongst related receptors from both Gram-positive and Gram-negative bacteria.

Addresses: [1]Biomolecular Research Institute, 343 Royal Parade, Parkville, Victoria 3052, Australia and [2]Molecular Microbiology Unit, Women's and Children's Hospital, North Adelaide, SA 5006, Australia.

*Corresponding author.
E-mail: mike.lawrence@bioresi.com.au

Key words: ABC-type binding protein, metal-binding protein, PsaA, *Streptococcus pneumoniae*

Received: 28 August 1998
Revisions requested: 23 September 1998
Revisions received: 5 October 1998
Accepted: 6 October 1998

Structure 15 December 1998, 6:1553–1561
http://biomednet.com/elecref/0969212600601553

Introduction

The pathogen *Streptococcus pneumoniae* is responsible for a number of serious diseases in humans, including pneumonia, meningitis and bacteraemia. Of particular clinical concern is the increasing level of resistance to penicillin and other anti-microbial drugs, due in large part to high usage of these agents. The emergence and spread of resistant strains has underscored the need to develop novel drug compounds and vaccines that are effective against the organism. The pneumococcal surface antigen PsaA has recently been proposed as a drug target [1]. $PsaA^-$ mutants have been shown to be avirulent in a mouse model and the mutant organism lacks the phenotypic ability to adhere to type II pneumocyte cell lines [2]. The molecule also shows negligible sequence variation in strains studied to date [3] and is a protective immunogen in the mouse — it is therefore also of clinical interest as a vaccine candidate [4].

Genomic sequence comparison with the broader streptococcal family suggests that PsaA belongs to an ATP-binding cassette-type (ABC-type) transport system [2]. ABC-type transport systems are found in both prokaryotic and eukaryotic organisms [5–7] and consist of up to three types of protein component: an extra-cytoplasmic protein responsible for solute binding, an integral membrane protein responsible for transport of the solute through the cell membrane, and a cytoplasmic protein that couples ATP hydrolysis to the transport process. The latter two components are readily identified in genomic sequencing by the presence of Walker-type motifs [8] within the ATP-binding protein and by the characteristic hydrophobicity profile of the integral membrane protein. The binding-protein component is always present in systems responsible for solute import but it is invariably absent from systems responsible for solute export [5]. In Gram-positive bacteria (such as *S. pneumoniae*) the binding proteins are extracellular and are presumed to be anchored to the cell membrane via a lipid that is covalently attached to the protein [7]. These lipoproteins have a characteristic LX_1X_2C motif (single-letter amino acid code, where X_1 is typically A, S, V, Q or T and X_2 is typically G or A [6]) at the C-terminal end of their signal sequence; the mature protein is presumed to form upon covalent attachment of a diacylglyceryl moiety to the cysteine followed by proteolytic cleavage, leaving the modified cysteine residue as the N terminus of the mature protein. By contrast, in Gram-negative bacteria the binding proteins lack a membrane anchor and are free in the periplasm. The importance of ABC-type systems in Gram-positive organisms has been highlighted by the release of the genomic sequence of *Bacillus subtilis* [9]: they form the most abundant class of protein in this organism. ABC-type binding proteins can be grouped into 'clusters' on the basis of their sequence relationships and ligand identity [6]; PsaA and a second pneumococcal ABC-type protein AdcA are

considered to belong to the newly defined 'Cluster 9' of metal transporters [10]. The requirements for growth of pneumococcal PsaA$^-$ and AdcA$^-$ mutants [10] suggest that these proteins play a role in the transport of Mn^{2+} and Zn^{2+} into the cytoplasm, respectively. However, there is no direct data on the relative affinities of PsaA and AdcA for these two metals.

Three-dimensional structures exist for a number of ABC-type binding proteins (for a review see [11]; more recent examples include the Fe^{+3}-binding protein from *Haemophilus influenzae* [12] and the molybdate-binding protein from *Escherichia coli* [13]). All known structural examples are from Gram-negative systems and none of these bears any recognizable sequence relationship to PsaA or is involved in the transport of Mn^{2+} or Zn^{2+}. In order to understand the function of PsaA in the pathogenesis of *S. pneumoniae* better, we have determined its three-dimensional structure to a resolution of 2.0 Å using X-ray crystallography. Our structure reveals an occupied metal-binding site within the molecule, with the ion most probably being Zn^{2+}. The structure also reveals a fundamental difference in the folding topology of this molecule in comparison to that of the broader family of ABC-type binding proteins, with implications for a novel mode of solute uptake and release.

Results

Overall structure of PsaA

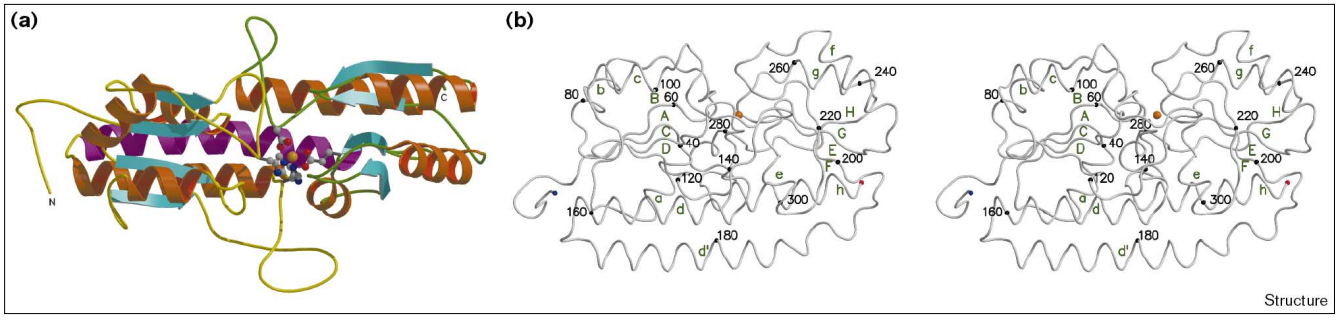
Our atomic model of PsaA has been refined to a crystallographic R factor of 0.178 (R_{free} = 0.236) and includes 280 residues of the 309-residue mature protein. PsaA consists of two $(\beta/\alpha)_4$ sandwich domains arranged in a twofold pseudosymmetric fashion (Figure 1a,b); we denote the N-terminal domain strands A–D and the helices a–d, whereas the C-terminal domain strands are denoted E–H and the helices e–h. The β strands within each domain form parallel sheets with a 2-1-3-4 linking topology. The domains are interlinked via a single helix (d′, residues 165–189) that acts as a 'backbone' to the structure, packing antiparallel to the fourth helix (d, residues 143–162) of the N-terminal domain and parallel to the fourth helix (h, residues 291–308) of the C-terminal domain. The location of the secondary structural elements within the PsaA sequence is shown in Figure 2. The domain interface also exhibits the twofold pseudosymmetry and is made up primarily of a juxtaposition of the first helix and second strand-to-helix loop of one domain with the fourth strand-to-helix loop of other domain.

The conformation of the N terminus of PsaA is not adequately defined by this structure. PsaA was crystallized [14] as an N-terminal His-tagged fusion protein with 12 additional residues preceding Cys20; in our crystals no interpretable electron density was obtained for residues preceding Lys24. The presence of the His tag precludes the covalent attachment of lipid to Cys20 in the *Escherichia coli* expression system. In our atomic model, residues 24–31 do not make extensive contact with the core elements of the N-terminal domain, and their conformation may in part be defined by a crystal contact in that region. Sequence analysis [2] reveals that these residues form the least-conserved region of the PsaA-related streptococcal proteins, and we therefore suggest that, *in vivo*, these residues form a somewhat disordered extension from the protein to the lipid surface.

Metal-binding site

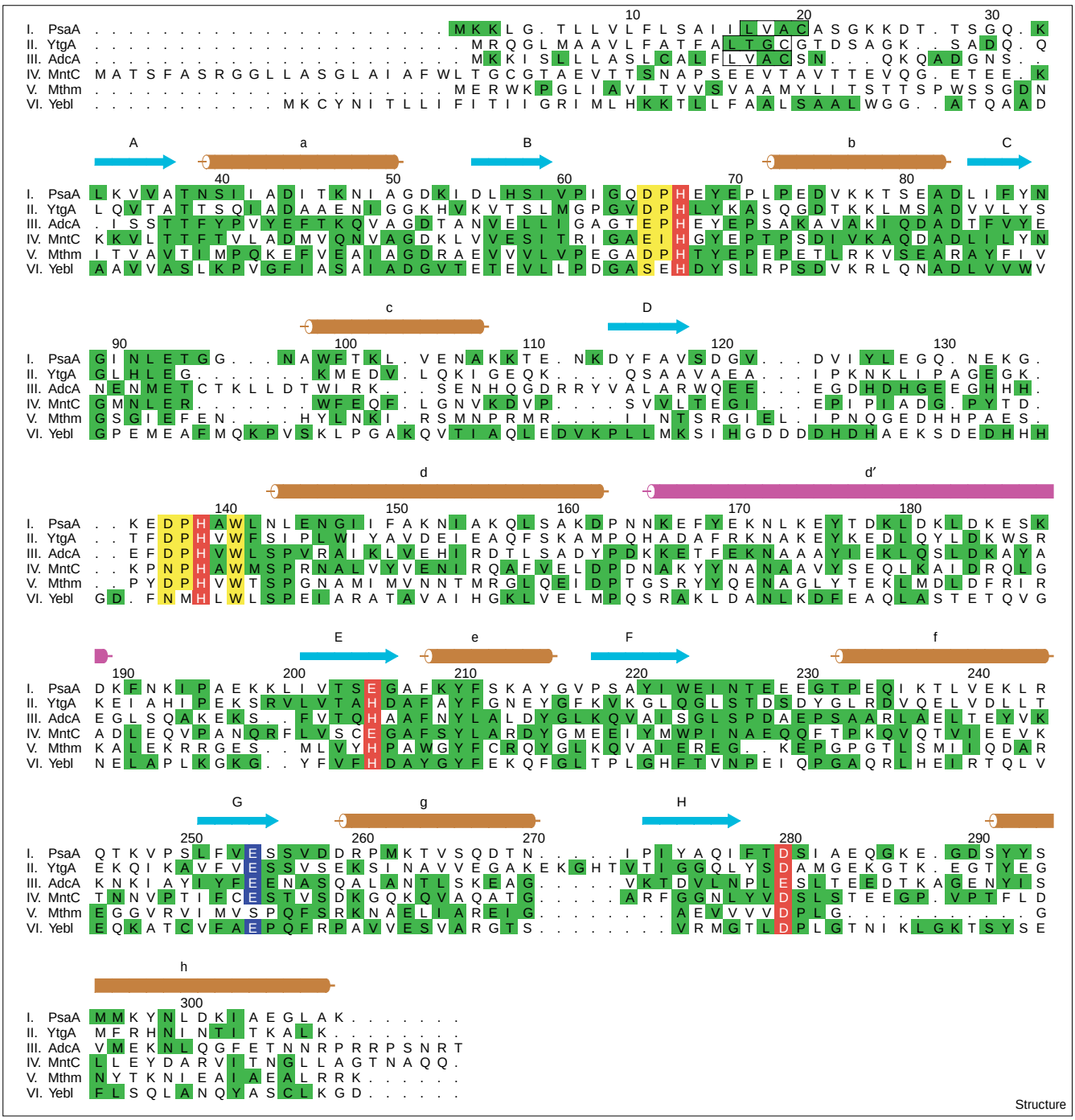
The metal-binding site identified in the electron-density maps obtained by multiple isomorphous replacement is situated in the interface between the N- and C-terminal

Figure 1



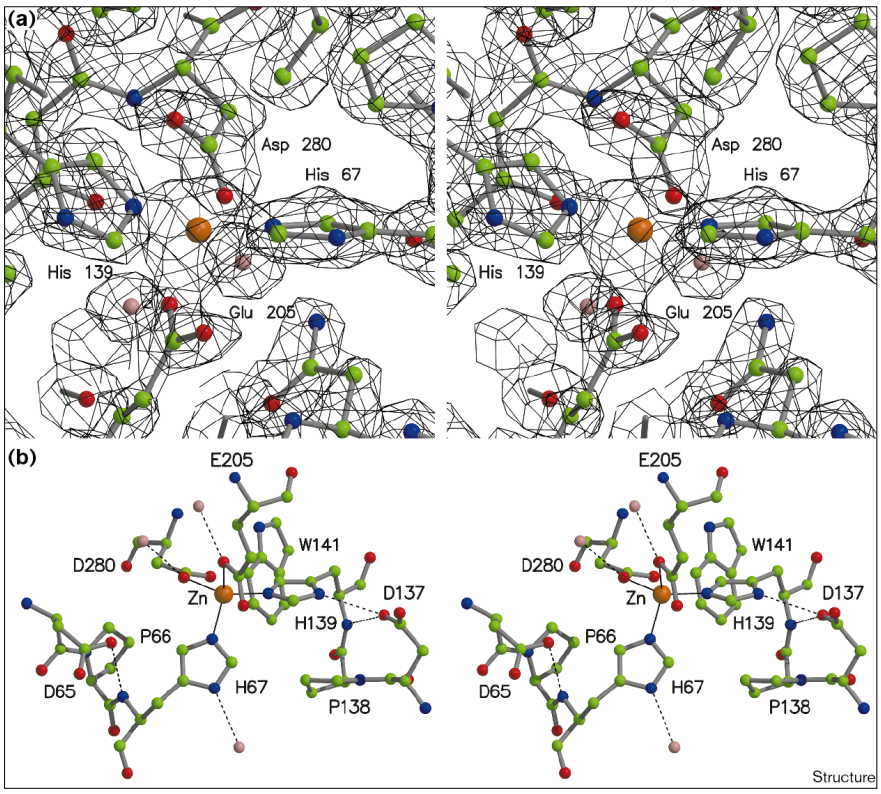
The overall fold of PsaA. (a) View down the twofold pseudosymmetry axis relating the domains. Within each domain, β strands are shown in cyan and α helices in brown; the domain-connecting helix is colored magenta. The connecting loops within the N-terminal domain are in yellow, whilst those in the C-terminal domain are colored green. The four metal-coordinating residues are shown in ball-and-stick representation, with carbon atoms colored gray, nitrogen atoms blue and oxygen atoms red. The Zn^{2+} is shown in orange. (b) Stereogram of a Cα trace of PsaA. The β strands are designated A–H and domain α helices a–h; the location of the Zn^{2+} (orange) is also shown. The N and C termini are shown as a blue and a red sphere, respectively, the observed N terminus being Lys24. Every twentieth Cα atom is highlighted as a black sphere and labeled. The view direction is approximately orthogonal to that of (a). Figures were generated using the programs MOLSCRIPT [39] and RASTER3D [40,41].

Figure 2



An alignment of sequences from Cluster 9 ABC-type binding proteins showing the relative locations of the PsaA secondary structural units and metal-binding site. The PsaA secondary structural elements are colored and labeled according to the scheme of Figure 1. Each subcluster I–VI is represented by a single member protein sequence and labeled accordingly (see Discussion section for definitions of subclusters and protein nomenclature). The numbering above the alignment is that of PsaA. Residues that are totally conserved within each subcluster are highlighted in green, and residues corresponding to the four primary PsaA metal-binding residues are highlighted in red; these are totally conserved within each subcluster. Trp141 and the first two residues of the DPH motifs are highlighted in yellow when totally conserved within each subcluster. The strand-bridging residue Glu254 is highlighted in blue when conserved within each subcluster. The LXXC lipoprotein motif is boxed in sequences from Gram-positive organisms. The sequence shown for PsaA is that of *S. pneumoniae* serotype 2 strain D39 [2], which, following this work, is now considered identical to that of serotype 6B strain SP-86 (JCP, unpublished results). The sequence of *S. pneumoniae* AdcA includes the correction Ile138 to His (JP Claverys, personal communication). Sequence alignment was performed using the program CLUSTALW [42] followed by some manual editing. The Figure was produced using the program ALSCRIPT 2.04 [43].

Figure 3



The PsaA metal-binding site. **(a)** Stereogram showing σ_A-weighted $(2F_o–F_c)$ electron density (contoured at the 1.5σ level) in the vicinity of the PsaA metal-binding site. The Zn^{2+} is shown in orange and the surrounding residues are in ball-and-stick representation (carbon atoms in green, oxygen in red and nitrogen in blue). Also shown (in pink) are the two water molecules hydrogen-bonded to Glu205 Oε1 and Asp280 Oδ2 (see text). **(b)** Stereogram showing the geometry of the PsaA metal-binding site. Atoms included are the four metal-coordinating residues, the DPH motifs and the water molecules associated with the metal-coordinating residues (see text). The color scheme is that of (a). The coordination geometry is highlighted by the solid black lines and potential hydrogen bonds are shown as dotted lines. The view direction is approximately that of Figure 1a. Figures were generated using MOLSCRIPT [39] and RASTER3D [40,41].

domains of the protein. Atomic modeling revealed that the site is formed by the sidechains of residues His67, His139, Glu205 and Asp280 (Figure 3a,b) and that it has a tetrahedral coordination geometry. His67 and His139 interact with the metal via their Nε2 nitrogen atoms whereas the carboxylate sidechains of Glu205 and Asp280 interact with the metal via their Oε1 and Oδ2 atoms, respectively. The metal-to-protein atomic distances are 1.99 Å for His67 Nε2, 2.01 Å for His139 Nε2, 2.04 Å for Glu205 Oε1 and 2.02 Å for Asp280 Oδ2, with the tetrahedral angles at the metal ion ranging from 93° to 131°. The metal-coordinating atoms Glu205 Oε1 and Asp280 Oδ2 are each hydrogen bonded to an ordered water molecule; these water molecules make no direct interaction with the metal, however. His67 and His139 form part of an Asp–Pro–His (DPH) motif partially conserved within Cluster 9 proteins (see Figure 2 and Discussion section). These DPH segments (residues 65–67 and 137–139) are closely similar to one another in conformation, apart from the opposite orientation of the respective imidazole rings (Figure 3b); allowing for the ring rotation, the nonhydrogen atoms of the two DPH motifs can be superimposed with a root mean square (rms) deviation of 0.79 Å. Asp137 Oδ1 is capable of forming a hydrogen bond with His139 Nδ1 (consistent with an increase in partial negative charge at the metal-coordinating His139 Nε1) and with His139 N. Although Asp65 Oδ1 also forms a hydrogen bond with His67 N, the opposite orientation of the imidazole ring of His67 implies that Asp65 Oδ1 is in van der Waals contact with His67 Cδ2. His67 Nσ1 is hydrogen bonded instead to a nearby buried water molecule. The metal ion is buried 5.7 Å beneath the molecular surface and the four coordinating residues are also entirely buried, with His139 and Asp280 being most distant from the molecular surface. The sidechains of His139 and Asp280 pack against that of Trp141; the latter residue is strictly conserved amongst the Cluster 9 proteins (Figure 2). It should be noted that, although the components of the domain interface have a pseudosymmetric arrangement, the four metal-coordinating residues are not symmetrically located on these elements.

Identity of the metal ion

During crystallographic refinement, the metal ion was modeled as Zn^{2+}; attempts to model it as Mn^{2+} failed — its temperature factor dropped to 4.5 $Å^2$ (compared to an average temperature factor of 8.6 $Å^2$ for the four coordinating atoms) with residual electron density still present at the 4.6 σ level at the metal site. In comparison, the temperature factor refined to 9.5 $Å^2$ for Zn^{2+} with residual electron density present at only the 2.6 σ level. In an attempt to confirm the choice of Zn^{2+} in preference to Mn^{2+}, an anomalous Fourier synthesis was computed using the protein phases and Bijvoet differences from the native data. The

Table 1

Data collection and MIR phasing statistics.

Data sets	Native 1	Native 2 / PIP*	$K_2Pt(NO_3)_4$	$HgAc_2$	$HgCl_2$	$UO_2(NO_3)_2$
No. of crystals	2	1	1	1	1	1
No. of reflections	7717	19,354	7871	10,039	7202	5928
Completeness % (outer shell)	95.7 (88.6)	99.5 (99.6)	96.2 (95.9)	92.0 (91.2)	99.7 (98.2)	100.0 (99.7)
$\langle I/\sigma(I)\rangle$ (outer shell)	11.2 (4.6)	12.4 (3.4)	6.0 (2.8)	9.4 (3.2)	11.9 (4.1)	11.5 (5.1)
Multiplicity (outer shell)	3.9 (3.3)	3.7 (3.5)	2.8 (2.7)	1.4 (1.4)	3.1 (2.8)	4.1 (4.0)
R_{sym}† (outer shell)	0.12 (0.27)	0.11 (0.39)	0.19 (0.38)	0.07 (0.24)	0.13 (0.34)	0.13 (0.28)
Resolution (Å)	2.7	2.0	2.7	2.4	2.8	3.0
Heavy atom (mM)		≪ 4	12	1.0	2.0	100
Soaking time‡		16 h	16 h	24 h	12 h	5 d
No. of sites		0	2	4	4	1
R_{iso}§		0.085 (to 3.0 Å)	0.202	0.163	0.172	0.188
Centric R_{cullis}¶			0.63	0.48	0.51	0.92
Acentric R_{cullis} (iso/ano)			0.64/0.94	0.43/0.82	0.46/0.93	0.89/0.97
Centric phasing power#			1.59	3.42	3.21	0.91
Acentric phasing power (iso/ano)			2.07/0.74	3.75/1.24	3.56/1.25	1.04/0.83
Overall FOM¥ centric/acentric (2.4 Å)	0.60/0.48					

*PIP: di-μ-iodo-bis(ethylene-diamine)-di-platinum(II)nitrate. †$R_{sym} = \Sigma_{hkl}(\Sigma_i(|I_{hkl,i}-\langle I_{hkl}\rangle|))/\Sigma_{hkl}\Sigma_i I_{hkl,i}$. ‡Soaking time: h, hours; d, days. §$R_{iso} = \Sigma(|F_{PH}-F_P|/\Sigma F_P$. ¶$R_{cullis}$ = ⟨phase-integrated lack of closure⟩ / $\langle|F_{PH}-F_P|\rangle$. #Phasing power = $\langle(F_H/$phase-integrated lack of closure$)\rangle$, see comments in Materials and methods section. ¥FOM = Figure of merit for centroid structure factor as defined in SHARP [26].

map yielded no significant density at the metal site. Although this could be interpreted as being consistent with the weaker anomalous signal of Zn^{2+} (f′ = 0.68 electrons) at CuK_α wavelength in comparison to that of Mn^{2+} (f′ = 2.81 electrons), the Bijvoet diffraction ratios (qf′) are only 0.003 and 0.012, respectively, and it is not clear that the quality of the native diffraction data (Table 1) would be sufficient to detect a single Mn^{2+} in this fashion.

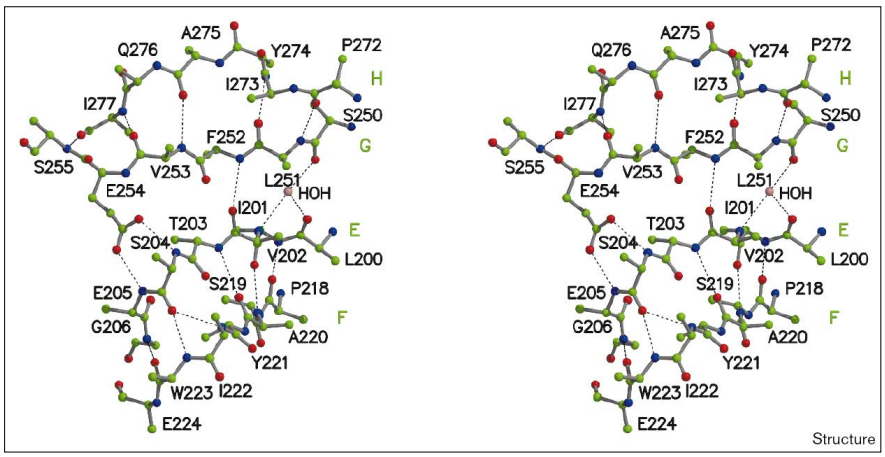
Architecture of the C-terminal domain

The association of β strands within the C-terminal domain is irregular (Figure 4). Only a single interbackbone hydrogen bond is found between strands E and G (Val202 O to Phe252 N). The interaction is mediated instead by the carboxylate oxygens of Glu254 (strand G), which hydrogen bond to the respective backbone nitrogen atoms of Ser204 and the metal-coordinating Glu205 (both in strand E), and by an ordered water molecule that hydrogen bonds to Val202 N (strand E), Ser250 O (strand G) and Leu200 O. These interactions prevent the formation of an intact β sheet between strands E and G by introducing both a shift in their relative register and additional twist in their relative rotation. The N-terminal residues of strand E (201–203) and strand F (218–220) associate via a canonical hydrogen-bond ladder. However, the ladder is broken by turns at residues 204 and 221, the ladder then resuming between residues 205/206 and 223. Five hydrogen bonds occur between the backbones of strands G and H, with a β bulge at Tyr274. These combined irregularities in the C-terminal strands contrast with the architecture of the N-terminal sheet, which has five canonical hydrogen bonds between strands A and B, four between A and C, and four between C and D. We note that the strand-bridging residue Glu254 is conserved in all Cluster 9 proteins shown in Figure 2 (with the exception of the more distantly related protein from *Methanobacterium thermoautotrophicum*), suggesting that this residue may play a critical role in maintaining the binding site geometry at Glu205. It may also play a role in structural aspects of metal uptake and release.

Comparisons with other structures

The structure of PsaA was compared to those in the Brookhaven Protein Data Bank (PDB) by submitting the Cα coordinates of PsaA to the DALI server [15]. The two most similar polypeptides were found to be the α and β subunits of the nitrogenase molybdenum-iron (MoFe) protein (PDB entries 3MIN and 2MIN, respectively; DALI Z scores 10.0 and 7.7, respectively), whereas the most similar ABC-type binding protein was found to be the L-arabinose-binding protein (PDB entry 8ABP; DALI Z score 4.9). DALI overlaid 206 Cα atoms of PsaA onto the β subunit of the nitrogenase MoFe protein [16] with an rms deviation of 3.6 Å, in comparison to 117 Cα atoms of PsaA onto the L-arabinose transporter [17] with an rms deviation of 4.1 Å. The nitrogenase MoFe protein is not an ABC-type binding protein, but is involved in the electron-transfer processes of dinitrogen reduction. Its very similar α and β subunits each consist of three domains; DALI finds PsaA to be similar in conformation to domains II and III of each of these subunits. Domain II of the β subunit has the same 2-1-3-4-linked parallel $(\beta/\alpha)_4$ topology observed in PsaA; domain III has a 3-2-1-4-5-linked $(\beta/\alpha)_5$ topology. Note that the 3-2-1-4-5 linking arrangement

Figure 4



Stereogram showing the irregular architecture of the β strands of the C-terminal domain of PsaA. The β strands are shown in ball-and-stick representation and are labeled as stated in the text. The sidechain of Glu254 is shown, but for all other residues sidechain atoms beyond Cβ are omitted. Also included is the ordered water molecule bridging strands E and G. The color scheme is that of Figure 3. Potential hydrogen bonds (computed using the program HBPLUS [44]) are shown as dashed lines. The Figure was generated using MOLSCRIPT [39] and RASTER3D [40,41].

is not fundamentally different from the 2-1-3-4 arrangement: it can be generated from the latter if a suitably long connecting peptide between the second and third β/α units is allowed to form an additional β/α unit appended to the sheet edge. Domains II and III have the same twofold pseudosymmetric arrangement that is observed in PsaA and are similarly connected by a single helix. The helix packs against the respective C-terminal helices of domains II and III in a fashion that is also similar to that observed in PsaA. The FeMo cofactor is located in the interface between domains II and III of the α subunit of the MoFe protein and hence may be considered to be in a structurally analogous location to that of the metal ion in PsaA. However, the FeMo cofactor is not as deeply buried in the domain II/III interface as the metal ion is in PsaA. The similarity of the L-arabinose-binding protein to PsaA is restricted to the limited aspects of the β/α architecture and to the relative spatial arrangement of the domains. The domains of the L-arabinose binding protein are connected by three extended polypeptide segments that add additional pairs of strands to the otherwise four-stranded sheets of its respective N- and C-terminal domains. It is in this aspect that the structure of PsaA is fundamentally different not only from that of the L-arabinose-binding protein but also from those of the broader family of ABC-type binding proteins (see Discussion section).

Discussion

Metal specificity

Experimental evidence suggests that the Cluster 9 proteins do not bind a common metal ion: PsaA is presumed to bind primarily Mn^{2+} and possibly Zn^{2+} [10], *S. pneumoniae* AdcA primarily Zn^{2+} [10], *Yersinia pestis* YfeA binds iron [18] and *Synechocystis* MntC binds Mn^{2+} [19]. We anticipate variation in metal specificity amongst these proteins to be reflected in variation in those residues that are sequence-related to the metal-coordinating residues of PsaA. To examine this hypothesis we have grouped Cluster 9 proteins into six subclusters (I–VI) on the basis of their overall primary-sequence identity. The subclusters contain the following proteins (GENBANK accession/locus numbers in parentheses) — I: *S. pneumoniae* PsaA (U53509), *S. sanguis* SsaB (M63481), *S. parasanguis* FimA (M26130), *S. gordonii challis* ScaA (L11577), *S. crista* ScbA (U46542), *Enterococcus faecalis* endocarditis-specific antigen (U03756), *Staphylococcus epidermidis* lipoprotein (X99127); II: *B. subtilis* YtgA (AF008220_22), *Erysipelothrix rhusiopathiae* EwlA (U52850), *Treponema pallidum* Tromp1 (U16363); III: *S. pneumoniae* AdcA (Z71552), *B. subtilis* YcdH (AB000617_20), IV: *Synechocystis* sp. MntC (L34630), *H. influenzae* HI0362 (U32720_9), *Y. pestis* YfeA (U50597); V: *Methanobacterium thermoautotrophicum* adhesion protein ('Mthm'; AE000842_8), *Synechocystis* sp. adhesion protein (D90903_59); VI: *E. coli* Yebl (D90828_10), *H. influenzae* HI0119 (U32698_2). The more distantly related periplasmic binding protein from *Archaeoglobus fulgidus* (AE000966_7) was not considered. Proteins within the same subcluster have a sequence identity of typically 35–95%, whereas the sequence identity between subclusters is typically 15–35%. It should be noted that, to date, a pair of Cluster 9 proteins has been identified in each of the organisms *S. pneumoniae, B. subtilis* and *H. influenzae* and that in each instance the members of the pair are located in different subclusters. A sequence alignment of representative members of each of the subclusters I–VI is shown in Figure 2. Inspection of the figure shows that His67 and His139 are strictly conserved across all six subclusters and although Glu205 and Asp280 are not strictly conserved across all subclusters, only a particular variant of these residues occurs within each subcluster (Glu205 is replaced by histidine in subclusters II, III, V and VI, and Asp280 is replaced by glutamate in subcluster III). We consider the DPH motifs to be an integral part of the metal-binding site: although this motif is not conserved

across all subclusters (Figure 2), again only a particular variant of the motif occurs within each subcluster. The only exception to this observation is in subcluster V. Hence, to a first approximation, each subcluster exhibits a distinct variant of the binding site. We therefore speculate that our subclustering partitions these proteins on the basis of the metal that they bind, or at least on the basis of their relative metal affinity. None of the coordinating residue alternatives observed across the subclusters would create a site theoretically incapable of zinc coordination [20] and therefore the variation in metal affinity and/or metal specificity would be dependent upon more subtle modulation of the site geometry or upon the inclusion of additional coordinating residues. Although we consider Zn^{2+} to be the metal ion in our X-ray structure, we cannot on the basis of this work exclude the affinity of PsaA for Mn^{2+} proposed by Dintilhac *et al.* [10]. However, we note that tetrahedral coordination of Mn^{2+} is unusual [21], but it is common for Zn^{2+} [20]. The affinity of PsaA for the observed ion is likely to be high because the protein was dialysed against an ostensibly metal-free solution prior to crystallization [14].

Metal binding and signal transduction

The interdomain helical 'backbone' of PsaA is novel amongst ABC-type receptors; this raises questions regarding the mode of PsaA metal binding and signal transduction. Where structures are known [11–13], Cluster 1–8 ABC-type binding proteins contain pairs of $\alpha/\beta/\alpha$ domains similar to, though often more elaborate than, those of PsaA, and their respective ligands range in size from ions to oligopeptides. In all these proteins (including the L-arabinose-binding protein highlighted in our DALI investigation) the polypeptide makes two or three interdomain crossovers, thereby adding additional β strands to the $(\beta/\alpha)_4$ domain cores. The crossover (or hinge) polypeptide segments in these proteins have an extended conformation and have been shown to allow movement of one domain with respect to the other [11], with the molecule capable of adopting a more open conformation when ligand-free. In Gram-negative bacteria the respective ABC-type permeases are believed to require such conformational change for recognition of the ligand-complexed binding protein [11]. It may be argued that there is no such requirement for conformational change in PsaA, given that its extracellular, membrane-anchored location could place it in permanent complex with the permease. However, if all periplasmic ABC-type binding proteins require conformational change for permease recognition, such an argument would appear to be at odds with the observation that there is no clear sequence distinction between periplasmic and extracellular Cluster 9 proteins apart from the LXXC motif. For example, the YtgA protein from the Gram-positive organism *B. subtilis* and the Tromp1 protein from the Gram-negative spirochaete *T. pallidum* (both in subcluster II) probably share a similar fold: they are 42% identical in sequence and are of comparable length. A lack of requirement for conformational change in Cluster 9 proteins also cannot be inferred solely from the small ionic nature of the liganded metal ions because hinge peptides have been found in the *H. influenzae* Fe^{3+}-binding protein [12]. Instead we suggest that conformational change does occur in Cluster 9 proteins, but that it is of a different structural nature to that hitherto observed. Cluster 9 would thus form an evolutionarily distant group of the ABC-type systems.

The peculiar asymmetry of the imidazole rings in the two DPH metal-binding motifs deserves further comment. It may be speculated that in the metal-free form both rings adopt the same conformation (i.e. with the imidazole Nδ2 atoms hydrogen bonded to the respective aspartic acid sidechain carboxylates). Capture of the metal would then, in part, involve a ring flipping of His167 in a trapdoor-like fashion. Such histidine imidazole ring-flipping has been observed in plastocyanin upon removal of copper [22].

Role of PsaA in virulence

Prior to their classification as ABC-type binding proteins, PsaA and its streptococcal homologs were considered to be adhesins [2]. However, the PsaA structure presented here is not reconcilable with an adhesin function if the molecule is indeed lipid-anchored to the bacterial membrane and encased beneath the peptidoglycan/teichoic-acid cell wall. The dimensions of PsaA are approximately 40 Å × 40 Å × 70 Å, with the putative lipid-modified N terminus situated at one end of the major axis. PsaA would then have no capacity to protrude through the cell wall and its surrounding polysaccharide capsule (a total distance of least 0.36 μm [23]). The loss of host adhesion observed in $PsaA^-$ mutants is thus more likely to be indirect, with a secondary protein being rendered absent or nonfunctional via Mn^{2+} (or Zn^{2+}) deprivation. Indeed, $PsaA^-$ mutant pneumococci have recently been reported not to express several surface proteins, including the choline-binding protein CbpA, a putative adhesin with specificity for receptors on cytokine-activated epithelial and endothelial cells (R Novak, E Charpentier, JS Braun, & EI Tuomanen, unpublished results). Intriguingly, the operon encoding the Cluster 9 ABC-type system ScaA in *S. gordonii challis* occurs adjacent to a gene for a hypothetical Zn-dependent endopeptidase [24], albeit on the opposite strand. The corresponding genomic region for *S. pneumoniae* has not yet been sequenced.

Our structure shows that the metal-binding site of PsaA is deeply buried in the domain interface and it is not immediately obvious how PsaA might be targeted with an antimicrobial agent. Two strategies present themselves. The first would be to target the metal-bound protein in order to inhibit metal release, as has been proposed for the *H. influenzae* Fe^{3+}-binding protein [12]. We note that in this protein, however, the Fe^{3+} lies considerably closer to

the surface than is the case in PsaA, and the coordinating residues are partly solvent exposed. The second possibility would be to target the putative 'open' form of the molecule, if it indeed exists. These strategies will need to be re-assessed when kinetic data of metal binding in PsaA become available.

Biological implications

Streptococcus pneumoniae (the pneumococcus) is a major human pathogen responsible for a variety of diseases including pneumonia, meningitis and bacteraemia. The pneumococcal surface protein PsaA is of interest both as a drug target and as a potential vaccine candidate. PsaA belongs to a family of so-called ABC-type (ATP-binding cassette-type) transport systems and is believed to be responsible for the uptake of Mn^{2+} and possibly Zn^{2+} as well. Although PsaA was initially considered to be an adhesin, its role in pathogenesis is now understood to be indirect: PsaA$^-$ mutants do not produce the putative adhesin CbpA. The structure of PsaA presented here reveals a metal-binding site located between two $(\beta/\alpha)_4$ domains in a fashion characteristic of all known ABC-type binding proteins. It is strikingly different from other known examples of these proteins, however, in that it lacks the so-called 'hinge peptides': the domains of PsaA are joined instead by a single helix that is in close association with the domain architecture. This suggests that PsaA has a novel mechanism of action and any conformational change is of a rather different nature from that observed until now. PsaA is a member of a broader 'cluster' of ABC-type metal-binding proteins. Identification of the metal-binding residues in PsaA provides a basis for interpreting the variation in metal specificity observed within this family, and also for suggesting the possible specificity of uncharacterized members of the cluster.

Materials and methods

The expression, purification, crystallization and native diffraction data collection of PsaA have been described [14]. Diffraction data from heavy-atom derivatives (Table 1) were collected under conditions identical to those described for the native crystals. Heavy-atom-site location was initially performed using the program RSPS from the CCP4 suite [25]. Heavy-atom refinement was then carried out using the program SHARP [26], with the $HgCl_2$ and $HgAc_2$ data sets being treated as two independent crystals from the same compound. Solvent flattening (using SOLOMON [27] within SHARP with a 50% solvent content) led to a readily interpretable map and the polypeptide could be traced unambiguously from residue 35 to 309. We note that SHARP produced phasing powers considerably higher than is typical for data of the quality presented. We were unable to resolve this issue but, given the ready interpretability of the map, we decided to proceed with atomic-model building. Map interpretation was assisted by a secondary structure model of PsaA obtained as an amalgam of results from the software packages PHD [28], PREDATOR [29] and DSC [30]. In retrospect, the prediction assigned 79% of the mature PsaA sequence to the correct secondary-structural class.

An initial atomic model was built using the program O [31], and improved by alternating rounds of simulated-annealing refinement within the program X-PLOR [32] and manual rebuilding. An intense feature (10.8 σ was present in the 2.4 Å solvent-flattened map and was modeled as a Zn^{2+} ion contributing only to the X-ray term of the refinement target function and not to the stereochemical restraints. The metal site was distinct from any site used in phasing (the latter sites being at least 19 Å away from the the metal site in the crystal). After the initial cycles of refinement, the native data set was replaced with a superior 2.0 Å data set obtained from a crystal soaked in di-μ-iodo-bis(ethylene-diamine)-di-platinum(II)nitrate (PIP) that showed little or no detectable heavy-metal substitution (Table 1). The final refinement steps employed REFMAC [33] and included a bulk-solvent correction and restrained individual B-factor refinement. Water molecules were added by alternating cycles of ARP [34] with REFMAC; final refinement statistics are given in Table 2. R_{free} values ([35], based on a test set of 5% of the reflections) were monitored throughout the refinement process and refinement steps were only accepted if they led to a reduction in R_{free}. The geometry of individual residues was monitored during refinement with OOPS [36] and any errors were corrected manually with O. The final model included residues 24–309, the metal ion and 225 solvent molecules. Ninety-one percent of the residues lay within the most favoured regions of the Ramachandran plot [37], and none lay in the disallowed regions. Residues 189, 192, 197 and 228–230 lay in poor electron density and were set to zero crystallographic occupancy during the final refinement. The sidechains of residues Glu68, Val217 and Ile277 were each visible in two conformations and these were built appropriately at half occupancy.

Table 2

Crystallographic refinement statistics.

Resolution range (Å)	8.0–2.0
Number of reflections	
working set	18,050
free set	979
R factor*	0.178
Free R factor	0.236
Rms deviation from ideality†	
bond distance (Å)	0.009
angle distance (Å)	0.026
planar 1–4 distance (Å)	0.028
Average B factors	
mainchain (Å^2)	15.0
sidechain (Å^2)	17.6
solvent (Å^2)	25.0
metal ion (Å^2)	9.5
Rms ΔB mainchain bond (Å^2)	1.2
Rms ΔB sidechain bond (Å^2)	1.6

*$R = \Sigma_{hkl} |F_o - F_c| / \Sigma_{hkl} F_o$. †Ideality as defined in the CCP4 v3.3 implementation of the Engh and Huber parameters [38].

Accession numbers

The coordinates have been deposited in the Brookhaven Protein Data Bank with accession code 1PSZ.

Acknowledgements

We thank Peter Colman for his support of this project.

References

1. Paton, J.C. (1998). Novel pneumococcal surface proteins: role in virulence and vaccine potential. *Trends Microbiol.* 6, 85-87.
2. Berry, A.M. & Paton, J.C. (1996). Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*. *Infect. Immun*. 64, 5255-5262.
3. Sampson, J.S., Furlow, Z., Whitney, A.M., Williams, D., Facklam, R. & Carlone, G.M. (1997). Limited diversity of *Streptococcus pneumoniae psaA* among pneumococcal vaccine serotypes. *Infect. Immun.* 65, 1967-1971.

4. Talkington, D.F., Brown, B.G., Tharpe, J.A., Koening, A. & Russell, H. (1996). Protection of mice against fatal pneumococcal challenge by immunization with pneumococcal surface adhesin A (PsaA). *Microb. Pathog.* 2, 17-22.
5. Higgins, C.F. (1992). ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* 8, 67-113.
6. Tam, R. & Saier, M.H., Jr. (1993). Structural, functional and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.* 57, 320-346.
7. Sutcliffe, I.C. & Russell, R.R.B. (1995). Lipoproteins of Gram-positive bacteria. *J. Bacteriol.* 177, 1123-1128.
8. Walker, J.E., Saraste, M., Runswick, M.J. & Gay, N.J. (1982). Distantly related sequences in the α- and β-subunits from ATP synthase, myosin, kinases and other ATP requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945-951.
9. Kunst, F., *et al.*, & Danchin, A. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 390, 249-256.
10. Dintilhac, A., Alloing, G., Granadel, C. & Claverys, J.-P. (1997). Competence and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Mol. Microbiol.* 25, 727-739.
11. Quiocho, F.A. & Ledvina, P.S. (1996). Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.* 20, 17-25.
12. Bruns, C.M., *et al.*, & McRee, D.E. (1997). Structure of *Haemophilus influenzae* Fe^{+3}-binding protein reveals convergent evolution within a superfamily. *Nat. Struct. Biol.* 4, 919-924.
13. Hu, Y., Rech, S., Gunsalus, R.P. & Rees, D.C. (1997). Crystal structure of the molybdate binding protein ModA. *Nat. Struct. Biol.* 4, 703-707.
14. Pilling, P.A., Lawrence, M.C., Berry, A.M., Ogunniyi, A.D., Lock, R.A. & Paton, J.C. (1998). Expression, purification and preliminary X-ray crystallographic analysis of PsaA, a putative metal-transporter protein of *Streptococcus pneumoniae*. *Acta Cryst . D* 54, 1464-1466.
15. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123-138.
16. Kim, J. & Rees, D.C. (1992). Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*. *Nature* 360, 553-560.
17. Vermersch, P.S., Lemon, D.D., Tesmer, J.J. & Quiocho, F.A. (1991). Sugar-binding and crystallographic studies of an arabinose-binding protein mutant (Met108Leu) that exhibits enhanced affinity and altered specificity. *Biochemistry* 30, 6861-6866.
18. Bearden, S.W., Staggs, T.M. & Perry, R.D. (1998). An ABC transporter system of *Yersinia pestis* allows utilization of chelated iron by *Escherichia coli* SAB11. *J. Bacteriol.* 180, 1135-1147.
19. Bartsevich, V.V. & Pakrasi, H.B. (1995). Molecular identification of an ABC transporter complex for manganese: analysis of a cyanobacterial mutant strain impaired in the photosynthetic oxygen evolution process. *EMBO J.* 14, 1845-1853.
20. Vallee, B.L. & Auld, D.S. (1995). Zinc metallochemistry in biochemistry. In *Interfaces between Chemistry and Biochemistry* (Jollès, P. & Jörnvall, H., eds), pp. 259-277, Birkhäuser-Verlag, Basel.
21. Christianson, D.W. (1997). Structural chemistry and biology of manganese metalloenzymes. *Prog. Biophys. Mol. Biol.* 67, 217-252.
22. Garrett, T.P.J., Clingeleffer, J., Guss, J.M., Rogers, S.J. & Freeman, H.C. (1984). The crystal structure of poplar apoplastocyanin at 1.8 Å resolution: the geometry of the copper-binding site is created by the polypeptide. *J. Biol. Chem.* 259, 2822-2825.
23. Tomasz, A. (1981). Surface components of *Streptococcus pneumoniae*. *Rev. Infect. Dis.* 3, 190-211.
24. Kolenbrander, P.E., Andersen, R.N. & Ganeshkumar, N. (1994). Nucleotide sequence of the *Streptococcus gordonii* PK488 coaggregation adhesin gene, scaA, and ATP-binding cassette. *Infect. Immun.* 62, 4469-4480.
25. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* 50, 760-763.
26. La Fortelle, E. de & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In *Methods in Enzymology* 276. (Sweet, R.M. & Carter, C.W. Jr, eds), pp. 472-494, Academic Press, New York.
27. Abrahams, J.P. & Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F_1 ATPase. *Acta Cryst. D* 52, 30-42.
28. Rost, B., Sander, C. & Schneider, R. (1994). PHD - an automatic mail server for protein secondary structure predictions. *Comput. Appl. Biosci.* 10, 53-60.
29. Frischman, D. & Argos, P. (1997). 75% accuracy in protein secondary structure prediction. *Proteins* 27, 329-335.
30. King, R.D. & Sternberg, M.J.E. (1996). Identification and application of concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci.* 5, 2298-2310.
31. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* 47, 110-119.
32. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* 235, 458-460.
33. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst. D* 53, 240-255.
34. Lamzin, V.S. & Wilson, K.S. (1993). Automated refinement of protein models. *Acta Cryst. D* 49, 129-147.
35. Brünger, A.T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472-474.
36. Kleywegt, G.J. & Jones, T.A. (1996). Efficient rebuilding of protein structures. *Acta Cryst. D* 52, 829-832.
37. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283-291.
38. Engh, R.A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray structure refinement. *Acta Cryst. A* 47, 392-400.
39. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* 24, 946-950.
40. Bacon, D. & Anderson, W.F. (1988). A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* 6, 219-220.
41. Merritt, E.A. & Murphy, M.E.P. (1994). Raster3D version 2.0 – a program for photorealistic molecular graphics. *Acta Cryst. D* 50, 869-873.
42. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acid Res.*, 22, 4673-4680.
43. Barton, G.J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* 6, 37-40.
44. McDonald, I.K. & Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238, 777-793.